



Algorithm for Punjabi Question Answering System

Poonam Gupta*

ME Research Scholar Computer Science & Engg.
University Institute of Engg and Tech.
Panjab University, Chandigarh, India

Vishal Gupta

Assistant Professor, Computer Science & Engg. Deptt.
University Institute of Engg. & Tech.
Panjab University, Chandigarh, India

Abstract— In this paper an algorithm for Punjabi Question Answering system has been implemented. The proposed system is designed and built in such a way that it increases the accuracy of Question Answering Systems in terms of Recall and Precision and is working for factoid questions and answers text in Punjabi. The system constructs a novel pattern finding and matching system to identify most accurate probable answer out of multiple answers. This algorithm is based on understanding the meaning of the given Punjabi question and expresses them in query logic language. The answers are extracted for the questions of type ਕੀ (what), ਕਦੇ (when), ਕਿੱਥੇ (where), ਕੌਣ (who) and ਕਿਉ (why).

Keywords— Natural Language Processing, Punjabi Question Answering System, Information retrieval.

I. INTRODUCTION

In today's context, there are three components of Question answering system (QAS) are question classification, information retrieval, and answer extraction. These components play a vital role in QAS. Question classification play chief role in QA system to categorize the question based upon its type. Information retrieval method is to identify success by extracting out applicable answer post by their intelligent question answering system [4]. In comparing with classical information retrieval, where whole documents are considered similar to the information demand, in question answering, specific pieces of information are returned as an answer. The user of a question answering system is interested in a concise, comprehensible, correct and most probable answer, which may refer to a word, sentence, paragraph, image, audio fragment, or an entire document [1]. Question answering system is an important application of the text mining [2]. These systems extract more relevant answers to a particular question posed to them from the large collection of text corpus [5]. On the other hand, various relevant answers can also be extracted classifying the questions into various categories and hence relevancy can be detected easily from the questions and the answers [6]. The Internet today has to face the difficulty of dealing with multi linguism [3]. All the work in Question answering system is done for various other languages but as per knowledge, limited work is done for Punjabi Language.

II. ALGORITHM FOR PUNJABI QUESTION ANSWERING SYSTEM

The proposed algorithm takes Punjabi question and a paragraph text as input from which answers are to be extracted. The algorithm proceeds by segmenting the Punjabi input question into Words. For each word in the question follow following steps:

Step 1 Query Processing: - In this step, it classify the questions based upon its categories like ਕੌਣ (WHO), ਕੀ (WHAT), ਕਿਉ (WHY), ਕਿੱਥੇ (WHERE), ਕਦੇ (WHEN) and the system checks the corresponding rules.

Step 2 If the input question consists of ਕੌਣ (WHO) then go to the procedure for ਕੌਣ (WHO).

elseIf the input question consists of ਕੀ (WHAT) then go to the procedure for ਕੀ (WHAT).

elseIf the input question consists of ਕਦੇ (WHEN) then go to the procedure for ਕਦੇ (WHEN).

elseIf the input question consists of ਕਿੱਥੇ (WHERE) then go to the procedure for ਕਿੱਥੇ (WHERE).

elseIf the input question consists of ਕਿਉ (WHY) then go to the procedure for ਕਿਉ (WHY).

Else Algorithm cannot find the answer.

Algorithm Input Example:

ਮਹਾਤਮਾ ਗਾਂਧੀ ਭਾਰਤ ਦੇ ਰਾਸ਼ਟਰ ਪਿਤਾ ਸਨ। ਆਪ ਅਹਿੰਸਾ ਦੇ ਅਵਤਾਰ ਸਨ। ਆਪ ਦਾ ਜਨਮ 2 ਅਕਤੂਬਰ, 1869 ਈ: ਨੂੰ ਪੇਰਬੰਦਰ (ਕਾਠੀਆਵਾੜ), ਗੁਜਰਾਤ ਵਿਚ ਹੋਇਆ। ਆਪ ਦਾ ਪੂਰਾ ਨਾਂ ਮੋਹਨ ਦਾਸ ਕਰਮ ਚੰਦ ਗਾਂਧੀ ਸੀ। ਆਪ ਸਦਾ ਸੱਚ ਬੋਲਦੇ ਸਨ ਤੇ ਮਾਤਾ - ਪਿਤਾ ਦੀ ਆਗਿਆ ਦਾ ਪਾਲਣ ਕਰਦੇ ਸਨ। ਦੇਸ਼ ਵਿੱਚ ਬੀ.ਏ ਪਾਸ ਕਰ ਕੇ ਵਲਾਇਤ ਤੋਂ ਬੈਰਿਸਟਰੀ ਦੀ ਡਿਗਰੀ ਪ੍ਰਾਪਤ ਕੀਤੀ। ਗਾਂਧੀ ਜੀ ਕਈ ਵਾਰ ਜੇਲ੍ਹ ਗਏ। 1931 ਈ: ਵਿੱਚ ਆਪ ਨੇ ਲੂਣ ਦਾ ਸਤਿਆਗ੍ਰਹ ਕੀਤਾ।

1942 ਈ: ਵਿਚ ਆਪ ਨੇ ਅੰਗਰੇਜ਼ਾਂ ਵਿਰੁੱਧ 'ਭਾਰਤ ਛੱਡੋ' ਲਹਿਰ ਚਲਾਈ। ਅੰਤ ਅੰਗਰੇਜ਼ਾਂ ਨੇ ਮਜਬੂਰ ਹੋ ਕੇ 15 ਅਗਸਤ, 1947 ਈ: ਨੂੰ ਹਥਿਆਰ ਸੁੱਟ ਦਿੱਤਾ। ਇਸ ਸਮੇਂ ਦੇਸ਼ ਦੀ ਵੰਡ ਦੇ ਕਾਰਨ ਹੋਏ ਫਿਰਕੂ ਫਸ਼ਾਦਾਂ ਨੂੰ ਦੇਖ ਆਪ ਬਹੁਤ ਦੁਖੀ ਹੋਏ। ਆਪ ਅਫ਼ਰੀਕਾ ਤੋਂ ਭਾਰਤ ਅੰਗਰੇਜ਼ਾਂ ਤੋਂ ਅਜ਼ਾਦ ਕਰਾਉਣ ਲਈ ਸੰਘਰਸ਼ ਕਰਨ ਵਾਪਸ ਆਏ। 30 ਜਨਵਰੀ, 1948 ਨੂੰ ਹਤਿਆਰੇ ਨੱਥੂ ਰਾਮ ਗੋਡਸੇ ਨੇ ਗੋਲੀਆਂ ਮਾਰ ਕੇ ਆਪ ਨੂੰ ਸਹੀਦ ਕਰ ਦਿੱਤਾ।

Questions asked from the above paragraph as below:

- 1) ਮਹਾਤਮਾ ਗਾਂਧੀ ਕੌਣ ਸਨ ?
- 2) ਮਹਾਤਮਾ ਗਾਂਧੀ ਦਾ ਜਨਮ ਕਦੋਂ ਅਤੇ ਕਿੱਥੇ ਹੋਇਆ ?
- 3) ਮਹਾਤਮਾ ਗਾਂਧੀ ਦਾ ਪੂਰਾ ਨਾਂ ਕੀ ਸੀ ?
- 4) ਮਹਾਤਮਾ ਗਾਂਧੀ ਅਫ਼ਰੀਕਾ ਤੋਂ ਭਾਰਤ ਵਾਪਸ ਕਿਉਂ ਆਏ ?

Output:-

ਮਹਾਤਮਾ ਗਾਂਧੀ ਭਾਰਤ ਦੇ ਰਾਸਟਰ ਪਿਤਾ ਸਨ।

ਆਪ ਦਾ ਜਨਮ 2 ਅਕਤੂਬਰ, 1869 ਈ: ਨੂੰ ਪੇਰਬੰਦਰ (ਕਾਠੀਆਵਾੜ), ਗੁਜਰਾਤ ਵਿਚ ਹੋਇਆ।

ਆਪ ਦਾ ਪੂਰਾ ਨਾਂ ਮੋਹਨ ਦਾਸ ਕਰਮ ਚੰਦ ਗਾਂਧੀ ਸੀ।

ਆਪ ਅਫ਼ਰੀਕਾ ਤੋਂ ਭਾਰਤ ਅੰਗਰੇਜ਼ਾਂ ਤੋਂ ਅਜ਼ਾਦ ਕਰਾਉਣ ਲਈ ਸੰਘਰਸ਼ ਕਰਨ ਵਾਪਸ ਆਏ।

A. Procedure for ਕੋਣ (WHO)

This procedure is used for finding the named entities from input paragraph. The algorithm for ਕੋਣ (Who) has been given below:

Step 1 Extract the substring before ਕੋਣ (WHO) from the given question and say it is S1.

Step 2 Extract the substring after ਕੋਣ (WHO) upto the end of the question and say it is S2.

Step 3 Initialize Answer = null.

Step 4 If S1 ≠ null then depending upon S1 set the answer

If S1 contains the name of the person

// Name is identified by searching current word in the Punjabi Dictionary if it is not in the Punjabi Dictionary and found in the Punjabi name list(it consists of 18000 names) then it is a name//

{

Search the paragraph by matching the strings in S1 and S2
and append the matching details to the end of S1. Say it
is S1'.

Append S2 after S1' and store it in the Answer.

}

} (End of Step 4 loop) else

Step 5 If S1 contains the words ਤੁਸੀਂ (tusī), ਤੁਹਾਨੂੰ (tuhānūṃ), ਤੂੰ (tūṃ) then

{

Replace that word as follows:

ਤੁਸੀਂ (tusī) to ਮੈਂ (maiṃ)

ਤੁਹਾਨੂੰ (tuhānūṃ) to ਮੈਨੂੰ (maiṃ)

ਤੂੰ (tūṃ) to ਮੈਂ (maiṃ).

}

Search the name of person, thing or concept in the paragraph
based upon S1.

//Name can be identified by searching in the Punjabi dictionary if it is not in the dictionary and it is found
in the Punjabi name list(it consists of 18000 names) then it is considered as a name// and append it after S1. Say it is
S1'.

Change the last word of S2 according to the words in S1.

{

If ਤੁਸੀਂ (tusī) then last word is ਹੋ (hō) changes to ਹਾਂ (hām).

If ਤੁਹਾਨੂੰ (tuhānūṃ) then last word is same as in substring S2. If ਤੂੰ (tūṃ) then last word is ਹੈ (hai) changes to ਹਾਂ
(hām) and store it in S2'.

Append S2' to the end of S1' and store it in Answer.

} (end of step 5 loop)

The procedure has got two rules out of which one must be satisfied to generate final result. If step4 of procedure satisfies the sentence then the question is of the form “ਤੁਸੀ ਕੋਣ ਹੈ? tusī kōṅ hō?” And the result generated contain pattern similar to the sentence “ਮੈਂ ਕਮਲਜੀਤ ਹਾਂ। mainṁ kamlajīt hām”. Similarly if step 4 satisfies the sentence then the question is of one of the two forms, it can be “ਤੁਹਾਨੂੰ ਕੋਣ ਜਾਣਦਾ ਹੈ? tuhānūṁ kōṅ jāṅdā hai?” And the result generated contain pattern similar to the sentence “ਮੈਨੂੰ ਮਨਦੀਪ ਜਾਣਦਾ ਹੈ। mainūṁ mandīp jāṅdā hai.” The second form is “ਤੂੰ ਕੋਣ ਹੈ? tūṁ kōṅ hai?(who are you?)” And the result generated contain pattern similar to the sentence “ਮੈਂ ਲਵਲੀਨ ਹਾਂ। “mainṁ lavlīn hām.”

B. Procedure for ਕੀ(WHAT)

This procedure is used for finding the reason for ਕੀ(WHAT) type of questions from the input paragraph. The algorithm for ਕੀ(WHAT) has been given below:

Step 1 Extract the substring before ਕੀ (WHAT) from the given question and say it is S1.

Step 2 Extract the substring after ਕੀ (WHAT) upto the end of the question and say it is S2.

Step 3 Initialize Answer = null.

Step 4 If S1= null then

```
{
    Store ਹਾਂ/ ਨਹੀ (Yes/No) in Answer by
    matching S2 with the paragraph.
} (end of step 4)
```

else

Step 5 If S1≠ null and contains the words ਤੁਸੀ (tusī) , ਤੁਹਾਡਾ (tuhāḍā), ਤੈਨੂੰ (tainūṁ),ਤੂੰ (tūṁ) then

```
{
    Replace the words as follows
    ਤੁਸੀ (tusī) to ਅਸੀ (asī)
    ਤੁਹਾਡਾ (tuhāḍā) to ਮੇਰਾ (mērā)
    ਤੈਨੂੰ (tainūṁ) to ਮੈਨੂੰ (mainūṁ)
    ਤੂੰ (tūṁ) to ਮੈਂ (mainṁ).
}
```

Search the name of person, thing or concept in the paragraph based upon S1

// name can be identified by searching in the Punjabi dictionary if it is not in the dictionary then it is considered as a name// Append it after S1. Say it is S1’.

Change the last word of S2 according to S1.

```
{
    If ਤੁਸੀ (tusī) then last word ਹੈ (hō) changes to ਹਾਂ (hām).
    If ਤੁਹਾਡਾ (tuhāḍā) then last word is same as in substring
    S2.
    If ਤੈਨੂੰ (tainūṁ) then last word is same as in
    substring S2.
    If ਤੂੰ (tūṁ) then last word is ਹੈ (hai) changes to ਹਾਂ (hām).
    Store it in S2’.
}
Append S2’ to S1’ and store it in answer.
} (end of step 5 loop)
```

This procedure has got two rules out of which one of the rule must satisfy to generate a final result. The Rule_1 is if the substring before ਕੀ (WHAT) is null then it stores ਹਾਂ/ ਨਹੀ (Yes/No) in Answer. The Rule_2 is if the substring before ਕੀ (WHAT) is not null then the question is of the form “ਤੁਸੀ ਕੀ ਖੇਡ ਰਹੇ ਹੈ? tusī kī khēḍ rahē hō?” And the result generated

contain the pattern “ਅਸੀ ਫੁੱਟਬਾਲ ਖੇਡ ਰਹੇ ਹਾਂ। asī phuṭṭbāl khēḍ rahē hām.” It can be “ਤੁਹਾਡਾ ਕੀ ਨਾਂ ਹੈ? tuhāḍā kī nām hai?” And the result generated contain the pattern “ਮੇਰਾ ਚਰਨਦੀਪ ਨਾਂ ਹੈ। mērā carandīp nām hai.” It can be “ਤੂੰ ਕੀ ਕਰ ਰਹੇ ਹੈ? tūm kī kar rahē hai?” And the result generated contain the pattern ਮੈਂ ਕੱਮ ਕਰ ਰਹੇ ਹਾਂ। maiṃ kamm kar rahē hām.

C. Procedure for ਕਦੋਂ (WHEN)

This procedure is used for finding the time and the date expression for a particular question from the input paragraph. The algorithm for ਕਦੋਂ (WHEN) has been given below:

Step 1 Extract the substring before ਕਦੋਂ (WHEN) from the given question and say it is S1.

Step 2 Extract the substring after ਕਦੋਂ (WHEN) upto the end of question and say it is S2.

Step 3 Initialize Answer = null

Step 4 If S1= null and S2 containing the words ਤੁਹਾਨੂੰ (tuhānūṃ), ਤੁਸੀ (tusī), ਤੁਹਾਡੀ (tuhāḍī), ਮੈਂ (maiṃ) then

{
Replace the word ਕਦੋਂ (WHEN) with the time and date expression and it can be identified if it is in the following formats:

DD/MM/YY like 02/04/97
DD/MM/YYYY like 02/04/1997
YY/MM/DD like 97/04/02
YYYY/MM/DD like 1997/04/02
MM/DD/YY like 05/26/97
MM/DD/YYYY like 05/26/1997
DD.MM.YY like 02.04.97
DD.MM.YYYY like 02.04.1997
DD-MM-YY like 02-04-97
DD-MM-YYYY like 02-04-1997
String, Numeric like May 26
Numeric, String like 26 May
Numeric : Numeric like 6:45 and
append ਨੂੰ and then store it in S1’.

Replace the words in S2

{
ਤੁਹਾਨੂੰ to ਮੈਨੂੰ (mainūṃ)
ਤੁਸੀ (tusī) to ਅਸੀ (asī)
ਤੁਹਾਡੀ (tuhāḍī) to ਮੇਰੀ
ਮੈਂ (maiṃ) to ਤੂੰ (tūṃ)
Store it in S2.

}
Append S2’ to the end of S1’ and store it in Answer.

}(end of step 4)

else

Step 5 If S1≠ null and is containing the word ਤੁਹਾਡੀ (tuhāḍī)

{
Then Replace the word ਤੁਹਾਡੀ (tuhāḍī) to ਮੇਰੀ. Store it in S1’.

Replace the word ਕਦੋਂ (WHEN) with the time and date expression and it can be identified if it is in the following formats

DD/MM/YY like 02/04/97
DD/MM/YYYY like 02/04/1997
YY/MM/DD like 97/04/02
YYYY/MM/DD like 1997/04/02
MM/DD/YY like 05/26/97
MM/DD/YYYY like 05/26/1997
DD.MM.YY like 02.04.97

DD.MM.YYYY like 02.04.1997
DD-MM-YY like 02-04-97
DD-MM-YYYY like 02-04-1997
String, Numeric like May 26
Numeric, String like 26 May

Numeric : Numeric like 6:45 and append ਨੂੰ and then store it in S1'.

Append S2 to the end of S1' and store it in Answer.
(end of step 5)

This procedure has one rule that must satisfy to generate a final result. The Rule is if the substring before ਕਦੇ (WHEN) is null then replace the word ਕਦੇ (WHEN) with the time and date expression. The question is of the form “ਕਦੇ ਤੁਸੀ ਇੱਥੇ ਤੋਂ ਜਾਵਾਰੋ? kadō tusī ithō tōm jāvāgē?” And the result generated contain the pattern “ਕਲ ਨੂੰ ਅਸੀ ਇੱਥੇ ਤੋਂ ਜਾਵਾਰੋ। Kal nūm asī ithō tōm jāvāgē.” The question can be of the form “ਕਦੇ ਤੁਹਾਨੂੰ ਇੱਥੇ ਪੰਜ ਸਾਲ ਹੋ ਜਾਵਾਰੋ? kadō tuhānūm ithhē pañj sāl hō jāvāgē?” And the result generated contain the pattern “02-04-2012, ਨੂੰ ਮੈਨੂੰ ਇੱਥੇ ਪੰਜ ਸਾਲ ਹੋ ਜਾਵਾਰੋ। 02-04-2012, nūm mainūm ithhē pañj sāl hō jāvāgē.” The question can be of the form “ਕਦੇ ਤੁਹਾਡੀ ਤਰੱਕੀ ਹੋ ਜਾਵੇਗੀ? kadō tuhāḍī tarkkī hō jāvēgī?” And the result generated contain the pattern 5-12- 2012 ਨੂੰ ਮੇਰੀ ਤਰੱਕੀ ਹੋ ਜਾਵੇਗੀ। 5-12-2012 nūm mērī tarkkī hō jāvēgī.” The question can be of the form “ਕਦੇ ਮੈਂ ਤੁਹਾਡੀ ਸਹਾਇਤਾ ਕਰ ਸਕਦਾ ਹਾਂ? kadō maim tuhāḍī sahāitā kar sakdā hām? And the result generated contain the pattern “04/12/2011 ਨੂੰ ਤੋਂ ਮੇਰੀ ਸਹਾਇਤਾ ਕਰ ਸਕਦਾ ਹਾਂ। 04/12/2011 nūm tōm mē sahāitā kar sakdā hām.”

D. Procedure for ਕਿੱਥੇ(WHERE)

This procedure is used for finding the location name for a particular question. The Algorithm for ਕਿੱਥੇ (WHERE) has been given below:

Step 1 Extract the substring before from the given question and say it is S1.

Step 2 Extract the substring after ਕਿੱਥੇ (WHERE) upto the end of question and say it is S2.

Step 3 Initialize Answer = null.

Step 4 If S1 ≠ null and S2 ≠ null and S1 is containing the words ਤੁਸੀ (tusī), ਤੁਹਾਨੂੰ (tuhānūm) then

```
{  
  Replace the words as follows  
  ਤੁਸੀ (tusī) to ਮੈਂ (maim)  
  ਤੁਹਾਨੂੰ (tuhānūm) to ਮੈਨੂੰ (mainūm) and store it in S1'.  
  Replace the word ਕਿੱਥੇ (WHERE) with the matching  
  locations like ਅੰਦਰ (In), ਤੇ (At), ਨੇੜੇ (Near) from  
  the paragraph or name of the location can also be  
  identified by searching it in the Punjabi dictionary  
  if it is not in the dictionary then it is considered as a  
  location name and store it in S1'.  
  Change the last word of S2 according to S1  
  {  
    If ਤੁਸੀ (tusī) then last word is ਹੋ changes to ਹਾਂ (hām).  
    If ਤੁਹਾਨੂੰ (tuhānūm) then last word is same as in  
    substring S2 and store it in S2'.  
  }  
}(end of step 4 loop)  
else
```

Step 5 If S1 is not containing the words ਤੁਸੀ (tusī), ਤੁਹਾਨੂੰ (tuhānūm) then

```
{  
  Keep S1 and S2 as such and store it in S1' and S2'  
  respectively.  
}
```

Just Replace ਕਿੱਥੇ (WHERE) with the matching locations

like ਅੰਦਰ (In), ਤੇ (At), ਨੇੜੇ (Near) from the paragraph or name of the location can also be identified by searching it in the Punjabi dictionary if it is not in the dictionary then it is considered as a location name and store it in S1'.

Append S2' to the end of S1' and store it in answer.

} (end of step 5 loop)

This procedure has one rule that must be satisfied to generate a final result. The Rule is if the substring before and after ਕਿੱਥੇ (WHERE) is not null then replace the word ਕਿੱਥੇ (WHERE) with the matching locations like ਅੰਦਰ (In), ਤੇ (At), ਨੇੜੇ (Near). The question is of the form “ਤੁਸੀਂ ਕਿੱਥੇ ਜਾ ਰਹੇ ਹੋ? tusī kaitthē jā rahē hō?” And the result generated contain the pattern “ਅਸੀਂ ਅੰਬਾਲਾ ਜਾ ਰਹੇ ਹਾਂ। asī ambālā jā rahē hām. It can be of the form “ਉਹ ਤੁਹਾਨੂੰ ਕਿੱਥੇ ਮਿਲੇਗਾ ? uh tuhānūṃ kaitthē mailēgā?” and the result generated contain the pattern “ਉਹ ਮੈਨੂੰ ਬੱਸ ਸਟੈਂਡ ਤੇ ਨੇੜੇ ਮਿਲੇਗਾ। uh mainūṃ bass saṭaiḍ tē nēṛē mailēgā.” It can also be of the form “ਇਹ ਸੱਜਨ ਕਿੱਥੇ ਦਾ ਰਹਿਣ ਵਾਲਾ ਹੈ? ih sajjan kaitthē dā rahaiṇ vālā hai?” and the result generated contain the pattern “ਇਹ ਸੱਜਨ ਦਿੱਲੀ ਦਾ ਰਹਿਣ ਵਾਲਾ ਹੈ। ih sajjan dailī dā rahaiṇ vālā hai.”

E. Procedure for ਕਿਉ (WHY)

This procedure is used for finding the reason for ਕਿਉ (WHY) type of questions from the input paragraph. The algorithm for ਕਿਉ (WHY) has been given below:

Step 1 Extract the substring ਕਿਉ (WHY) before from the given question and say it is S1.

Step 2 Extract the substring after ਕਿਉ (WHY) upto the end of question and say it is S2.

Step 3 Initialize Answer = null.

Step 4 If S1 ≠ null and S2 ≠ null then

{
 Keep S1 as such.
 Replace the word ਕਿਉ (WHY) with some reason by
 matching the strings in S1 and S2 with the paragraph.
 Store it after S1.
 Keep S2 as such.
 Append S2 to the end of S1 and store it in the Answer.

} (end of step 4 loop)

else

Step 5 If S1 ≠ null , S2 ≠ null and S1 is containing the words ਤੁਹਾਡੇ (tuhāḍē), ਤੁਸੀਂ (tusī)

{
 Replace the words as follows

{
 ਤੁਹਾਡੇ (tuhāḍē) to ਮੇਰੇ (mērē)

 ਤੁਸੀਂ (tusī) to ਅਸੀਂ (asī)

 }

 Replace the word ਕਿਉ (WHY) with some reason by

 matching the strings in S1 and S2 with the paragraph. Store
 it in S1.

 Keep S2 as such.

 Append S2 to the end of S1 and store it in the Answer.

} (end of step 5 loop)

This procedure has one rule that must be satisfied to generate a final result. The Rule is if the substring before and after ਕਿਉ (WHY) is not null then replace the word ਕਿਉ (WHY) with some reason by matching the substrings of the question. The question is of the form “ਬੱਚੇ ਜਮਾਤ ਵਿੱਚ ਖੁਸ਼ ਕਿਉ ਲਗ ਰਹੇ ਸਨ? baccē jamāt vaicc khush kaiu lag rahē san?” and the result generated contain the pattern “ਬੱਚੇ ਜਮਾਤ ਵਿੱਚ ਖੁਸ਼ ਕਮਲ ਦੇ ਜਨਮਦਨ ਤੇ ਲਗ ਰਹੇ ਸਨ। baccē jamāt vaicc khush kamal dē janmadain tē lag rahē san.

III. DATASET

TABLE I
DATASET

Total no. of Documents	50
Total Words	19410
Total Sentences	4800
Total characters	40
No. of questions asked	4850

$$\text{Accuracy} = \text{No. of correct answers} / \text{Total no. of questions asked} \times 100$$

$$= 3550/4850 = 73\%$$

TABLE II
ACCURACY TABLE

Accuracy of Punjabi Question Answering Algorithm for Document 1	50%
Accuracy of Punjabi Question Answering Algorithm for Document 2	100%
Accuracy of Punjabi Question Answering Algorithm for Document 3	100%
⋮	
Accuracy of Punjabi Question Answering Algorithm for Document 50	85%

$$\text{Average Accuracy} = \sum (\text{Accuracy of Punjabi Question Answering Algorithm for all the Documents} / 50) \times 100$$

$$= 3664/50 = 73\%$$

27% errors are due to these types of questions:

ਤੁਸੀਂ ਆਪਣੇ ਪੇਪਰ ਦੀ ਤਿਆਰੀ ਕਿਵੇਂ ਕੀਤੀ? *tusī āpaṇē pēpar dī tiārī kivēm kītī ?*

ਬਿਜਲੀ ਦੇ ਕਾਰਣ ਕਿਹੜੀ-ਕਿਹੜੀ ਸਮੱਸਿਆਵਾਂ ਖੜੀਆਂ ਹੁਈ ? *bijlī dē kāraṇ kihṛī-kihṛī samssiāvām kharāṁ huī ?*

ਪੰਜਵੀਂ ਬਰਸੀ ਕਿਸ ਦੀ ਸੀ ? *pañjvī barsī kis dī sī ?*

ਕਿੰਨੀ ਬੋਤਲਾਂ ਬਰਾਮਦ ਹੋਈ? *kinnī bōtalām barāmad hōī?*

ਮਾਮਲਾ ਕਿੰਨੇ ਵਿਅਕਤੀਆਂ ਵਾਸਤੇ ਦਰਜ ਕੀਤਾ ਗਿਆ ? *māmlā kinnē viaktīām vāsātē daraj kītā giā ?*

Since we did n't have the question types ਕਿਵੇਂ, ਕਿਹੜੀ-ਕਿਹੜੀ, ਕਿਸ, ਕਿੰਨੀ, ਕਿੰਨੇ that's why our algorithm doesn't return right answer for these type of questions. In future we will also implement these question types.

IV. CONCLUSIONS AND FUTURE WORK

The work illustrated here in context of Punjabi Question Answering System is first of its kind. In this paper we have discussed an algorithm for the implementation of Punjabi Question Answering System. Most of the resources used in this implementation are Punjabi Standard abridged dictionary and Punjabi name list. As no work has been done previously in this direction. The focus of the system has been basically on four kind of questions of type ਕੀ (what), ਕਦੇ (when), ਕਿੱਥੇ (where), ਕੋਣ (who) and ਕਿਉਂ (why). This system can be helpful for developing other NLP applications. For future scope, since in the current research work, we have used the concept making a hybrid that works in pattern matching (regular expressions) and new proposed answer finding scoring system, which has yielded for better Recall and Precision value, however for future scope we suggest more work must be done to add more types of questions for question classification and more types of answer patterns must be incorporated with scoring system that balances the probability between different natures of answers.

ACKNOWLEDGMENT

Many thanks to Mr. Vishal Gupta, Assistant Professor in UIET, Panjab University Chandigarh for doing this work.

REFERENCES

- [1] Oleksandr Kolomiyets and Marie –Francine Moens, “ A survey on question answering technology from an information retrieval perspective,” Elsevier J. Information Sciences, 181, pp. 5412-5434, 2011.
- [2] Vishal Gupta, and Gurpreet S. Lehal, “ A survey of text mining techniques ,” J. Emerging Tech. in Web Intelligence, vol.1 no.1, August 2009.
- [3] Shriya Sahu, Nandkishor Vasnik, and Devshri Roy, “Prashnottar : A Hindi question answering system,” International J. CSE and IT, vol.4 no.2, April 2012.
- [4] Muthukrishan Ramprasath and Shanmugaandram Hariharan, “A survey on question Answering System,” International J. Research and Reviews in Information Sciences, vol.2 no.1, March 2012.
- [5] Mohammad Reza Kangavari, Samira Ghandchi and Manak Golpour “A New Model for Question Answering Systems”, Journal of World Academy of Science, Engineering and Technology 42, pp. 506-513, 2008.
- [6] Liang Zhenqiu, “Design of Automatic Question Answering base on CBR”, Journal of Procedia Engineering 29, pp. 981-985, 2011.